MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 8: 121–143 (2024)
DOI: 10.3897/mbmg.8.104655

Research Article

Optimising species detection probability and sampling effort in lake fish eDNA surveys

Graham S. Sellers[1], Christopher L. Jerde[2], Lynsey R. Harper[1,3], Marco Benucci[1,4], Cristina Di Muri[1,5], Jianlong Li[6], Graeme Peirson[7], Kerry Walsh[7], Tristan Hatton-Ellis[8], Willie Duncan[9], Alistair Duguid[9], Dave Ottewell[10], Nigel Willby[11], Alan Law[11], Colin W. Bean[12], Ian J. Winfield[13], Daniel S. Read[14], Lori Lawson Handley[1], Bernd Hänfling[1,15]

1 *School of Natural Sciences, University of Hull, Hull, UK*
2 *Marine Science Institute, University of California Santa Barbara, Santa Barbara, CA, USA*
3 *The Freshwater Biological Association, The Hedley Wing, YMCA North Campus, Lakeside, Newby Bridge, Cumbria, LA12 8BD, UK*
4 *Fera Science Ltd, York Biotech Campus, Sand Hutton, York, UK*
5 *National Research Council, Research Institute on Terrestrial Ecosystems (CNR, IRET), University of Salento, Lecce, 73100, Italy*
6 *State Key Laboratory of Marine Resources Utilization in South China Sea, Hainan University, Haikou, Hainan, China*
7 *Environment Agency, Horizon House, Deanery Road, Bristol, BS1 5AH, UK*
8 *Natural Resources Wales, Maes Newydd, Britannic Way, Llandarcy, Neath SA10 6JQ, UK*
9 *Scottish Environment Protection Agency, Strathallan House, Castle Business Park, Stirling, UK*
10 *Natural England, Foss House, Kings Pool, Peasholme Green, York, YO 7PX, UK*
11 *Biological and Environmental Sciences, University of Stirling, Stirling, UK*
12 *NatureScot, Caspian House, Mariner Court, Clydebank Business Park, Clydebank, UK*
13 *Lake Ecosystems Group, UK Centre for Ecology & Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancaster, UK*
14 *UK Centre for Ecology & Hydrology (UKCEH), Wallingford, UK*
15 *UHI Inverness, Inverness, UK*
Corresponding author: Graham S. Sellers (graham.s.sellers@gmail.com)

OPEN ACCESS

Academic editor: Bettina Thalinger
Received: 6 April 2023
Accepted: 18 April 2024
Published: 24 July 2024

Citation: Sellers GS, Jerde CL, Harper LR, Benucci M, Di Muri C, Li J, Peirson G, Walsh K, Hatton-Ellis T, Duncan W, Duguid A, Ottewell D, Willby N, Law A, Bean CW, Winfield IJ, Read DS, Handley LL, Hänfling B (2024) Optimising species detection probability and sampling effort in lake fish eDNA surveys. Metabarcoding and Metagenomics 8: e104655. https://doi.org/10.3897/mbmg.8.104655

Copyright: © Graham S. Sellers et al.
This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

Environmental DNA (eDNA) metabarcoding is transforming biodiversity monitoring in aquatic environments. Such an approach has been developed and deployed for monitoring lake fish communities in Great Britain, where the method has repeatedly shown a comparable or better performance than conventional approaches. Previous analyses indicated that 20 water samples per lake are sufficient to reliably estimate fish species richness, but it is unclear how reduced eDNA sampling effort affects richness, or other biodiversity estimates and metrics. As the number of samples strongly influences the cost of monitoring programmes, it is essential that sampling effort is optimised for a specific monitoring objective. The aim of this project was to explore the effect of reduced eDNA sampling effort on biodiversity metrics (namely species richness and community composition) using algorithmic and statistical resampling techniques of a data set from 101 lakes, covering a wide spectrum of lake types and ecological quality. The results showed that reliable estimation of lake fish species richness could, in fact, usually be achieved with a much lower number of samples. For example, in almost 90% of lakes, 95% of complete fish richness could be detected with only 10 water samples, regardless of lake area. Similarly, other measures of alpha and beta-diversity were not greatly affected by a reduction in sample size from 20 to 10 samples. We also found that there is no significant difference in detected species richness between shoreline and offshore sampling transects, allowing for simplified field logistics. This could potentially allow the effective sampling of a larger number of lakes within a given monitoring

budget. However, rare species were more often missed with fewer samples, with potential implications for monitoring of invasive or endangered species. These results should inform the design of eDNA sampling strategies, so that these can be optimised to achieve specific monitoring goals.

Key words: eDNA metabarcoding, meta-analysis, sampling effort, species detection

Introduction

Environmental DNA (eDNA) metabarcoding of water samples is now regularly used for the detection and monitoring of fish species and the assessment of fish community structure (Wang et al. 2021). It is a non-invasive method which can be more effective at detecting elusive species than established invasive surveying techniques such as electrofishing, fyke netting or gill netting (Hänfling et al. 2016a; Pont et al. 2018; Lawson Handley et al. 2019; Griffiths et al. 2020; McElroy et al. 2020; Czeglédi et al. 2021; Pukk et al. 2021). Aquatic eDNA metabarcoding relies on the capture, extraction and sequencing of DNA within a water sample from a water body or a watercourse. However, DNA is rarely homogeneously distributed in aquatic environments (Beentjes et al. 2019; Lawson Handley et al. 2019; Bedwell and Goldberg 2020; Pukk et al. 2021). This is especially true in lentic environments where the dispersion of eDNA through hydraulic processes is often limited compared to lotic or marine environments Harrison et al. (2019). Caging experiments have shown that fish detection probability declines strongly within metres from the source in ponds (Li et al. 2019b; Brys et al. 2021) and tens of metres in lakes Dunker et al. (2016). Hence fish species detection relies on the collection of an adequate number of samples from a water body and their appropriate spatial distribution to capture the heterogeneity of the eDNA signal (Bruce et al. 2021). Multiple environmental processes affect the dispersion and degradation of eDNA in aquatic ecosystems including biotic mechanisms such as microbial communities and abiotic mechanisms such as ph, temperature and hydrology (Barnes and Turner 2015; Harrison et al. 2019). The presence of thermal stratification has been shown to lead to a more localised distribution of eDNA in lake ecosystems compared to well-mixed conditions with implications for the design of effective sampling strategies as (Lawson Handley et al. 2019; Hervé et al 2022). Because of the seasonal variation in lake stratification and activity of individual species, the timing of sampling is an important consideration for sampling in lentic systems (Hayami et al. 2020). Sampling strategies also vary according to the research question and are generally more intensive for detection of rare and/or low abundance species (Jerde et al. 2011; Dejean et al. 2012; Piggott et al. 2021) and determining fish species richness in high diversity ecosystems (Cantera et al. 2019; Blackman et al. 2021), than when the requirement is simply to establish the presence of common, widely distributed species (Sato et al. 2017). It is therefore important to determine the minimum number of samples required to achieve a specific outcome as cost effectiveness is essential for most biomonitoring programmes (Milián-García et al. 2021).

In this context, the UK Technical Advisory Group (UKTAG) on the European Union Water Framework Directive (WFD) initiated a research programme to

evaluate the suitability of eDNA metabarcoding approaches for monitoring lake fish communities, largely with the objective to develop a tool which is compatible with requirements under the WFD, i.e. to assess the ecological status of lakes. The research output of the original pilot study was published in 2016 (Hänfling et al. 2016a), with subsequent development of the method published in Li et al. (2018), Sellers et al. (2018) and Lawson Handley et al. (2019). The findings of this pilot demonstrated that 20 water samples were sufficient to detect the vast majority of fish species from England's largest lake, Windermere, and to provide ecologically meaningful relative abundance estimates (Hänfling et al. 2016). Subsequent results indicated that maximum species richness could be achieved by simply collecting samples from the shoreline during winter, likely due to increased water mixing as a result of more turbulent conditions (e.g. greater rainfall and winds) and less thermal stratification (Lawson Handley et al. 2019). A recent comprehensive literature review concluded that there is still insufficient knowledge about the required extent of spatial sampling to efficiently characterise fish communities in lentic systems to provide clear guidance on sampling strategies (Yao et al. 2022).

The primary aim of this study was to carry out a comprehensive analysis of an eDNA metabarcoding data set encompassing 101 lakes representing a range of lake types and environments across Great Britain. Our objective was to investigate how sample quantity and sampling location affect the estimation of fish biodiversity metrics, specifically species richness, community composition using both random and non-random data resampling techniques. To date, the number of samples necessary to achieve a 95% coverage threshold of the total species detected has received limited attention, but this is a critical aspect for optimising the cost-effectiveness of monitoring programmes. Based on the normal asymptotic shape of species accumulation curves, we hypothesise that a reduction in the number of water samples from the original data set will still be adequate to detect most fish species in any given UK lake, regardless of its area. We further hypothesise, based on our previous study, that biodiversity metrics obtained from shoreline and offshore samples do not differ significantly within lakes.

Methods

Study lakes and water sample collection

We utilised eDNA metabarcoding data from 101 lakes which were sampled between January 2015 and March 2019 largely during the winter season (November - March, Fig. 1). This includes previously published data from 14 Cheshire Meres and Welsh lakes (Li et al. 2019a). Lakes were chosen to represent various typologies (UKTAG 2004) representative across Great Britain, including alkalinity and ecological quality (Fig. 1). The surface area spectrum ranged from Scoat Tarn (4.3 ha) to Great Britain's largest, Loch Lomond (5158.7 ha), and included shallow lowland lakes as well as deep upland lakes. Even the northernmost British lakes including those sampled in this study are rarely covered by winter ice and typically do not exhibit significant winter stratification. A pre-existing classification of lake quality based on fully intercalibrated methodologies for assessing ecological status according to the EU Water

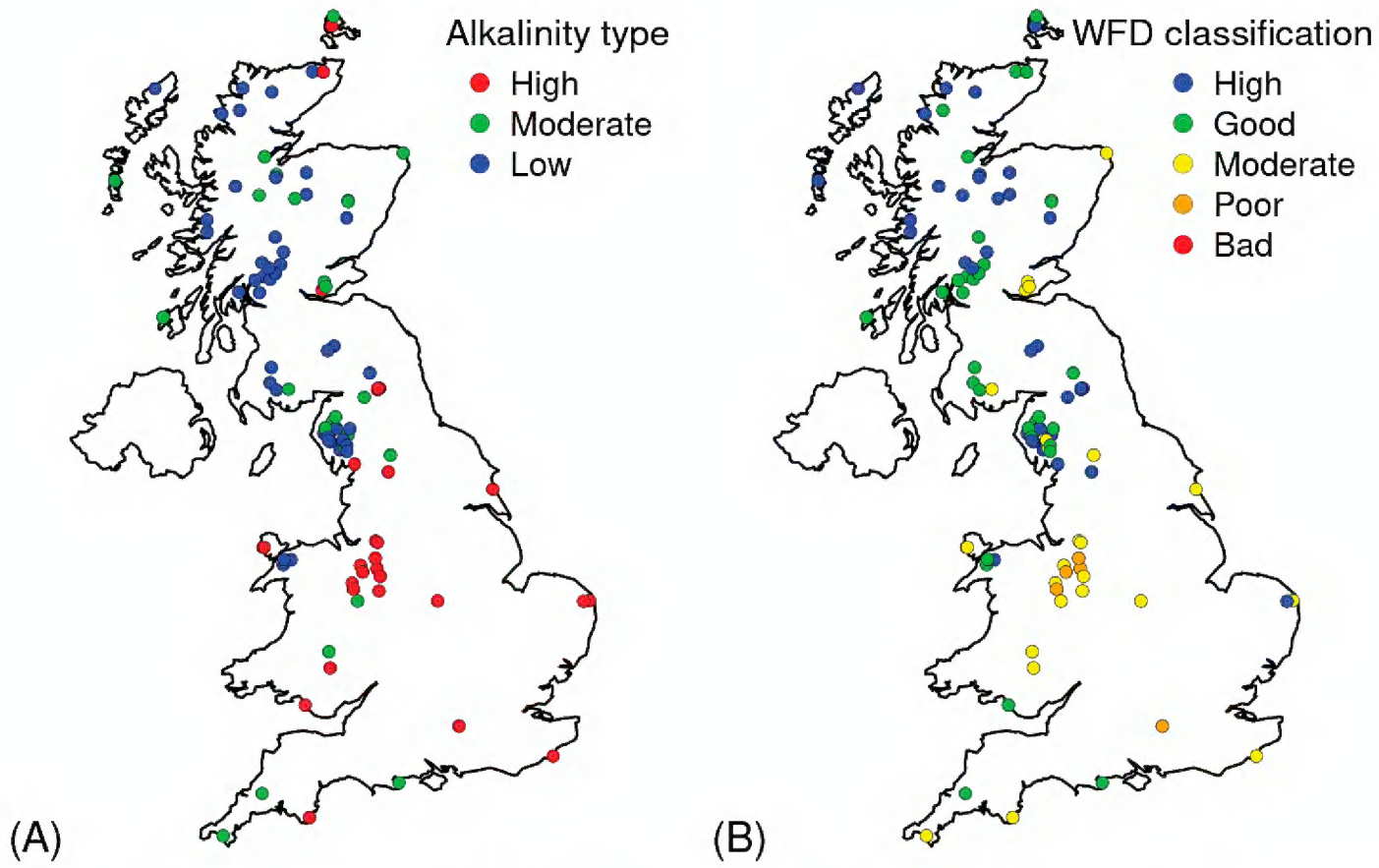

Figure 1. Distribution and characteristics of 101 UK lakes sampled for eDNA in this study. Shown are alkalinity type (left) and existing EU Water Framework Directive (WFD) classification (right) for each lake that takes account of Total Phosphorus, phytoplankton, diatoms, macrophyte and littoral invertebrates. For alkalinity types: High = >50 mg/L $CaCO_3$; Medium = 10–50 mg/L $CaCO_3$; Low =<10 mg/L $CaCO_3$. WFD classifications are based on an aggregate view of data for biological and physicochemical quality elements collected over the previous five years. Reproduced based on data from Willby et al. (2019).

Framework Directive (Birk et al. 2013) was available for all lakes which integrated the official classifications reported since 2009 for Total Phosphorus, phytoplankton, macrophytes, diatoms and littoral invertebrates (Fig. 1B). A consistent approach was used for sample collection and filtration as described in Hänfling et al. (2016b, 2016c; Hänfling et al. 2020). Shoreline samples were collected from all 101 lakes. Each individual shoreline sample contained 2 L of surface water and was composed of subsamples from five points along a 100 m transect, parallel to the shoreline. Where possible, 20 shoreline samples were collected at roughly equidistant points around the perimeter of each lake. Due to logistic constraints and varying objectives during early project phases, the actual number of shoreline samples collected across all lakes ranged from 10 to 21 shoreline samples (mean 17.74 ± 4.01 SD). An additional 8 to 25 offshore samples (mean 14.10 ± 5.67 SD) were collected from 20 of the lakes using a Friedinger or Ruttner sampler deployed at a specified depth. Each 2 L offshore sample was a composite of 5 × 400 mL samples collected from five points within a radius of 100 m around the sampling point. Each subsample was collected at a different depth covering the entire water column from surface to 1 m above the lake bottom. At least one field blank was included for each lake. A 2 L bottle containing purified water was carried alongside water sampling and stored with the samples during transport.

Water filtration and DNA extraction

Samples were stored immediately in cool boxes on ice, and filtered within 24 hours of collection. Samples were vacuum filtered through sterile Whatman 0.45 µm 47 mm cellulose nitrate membrane or mixed cellulose ester filters

(GE Healthcare). Two litres were filtered when possible, but filtration time was capped at one hour. Two filters were used for turbid samples, and later combined in a single DNA extraction step. Filters for each sample were stored separately at -20 °C until extraction.

Two slightly different but related protocols were used for DNA extraction over the course of the project. During the initial phase (2015–2017; *n* = 20 lakes; Hänfling et al. 2016a; Li et al. 2019a; Lawson Handley et al. 2019), DNA was extracted from filters using the MoBio PowerWater DNA Isolation Kit (now Qiagen DNeasy PowerWater Kit). In later phases (2017 - present, *n* = 81 lakes), DNA was extracted from filters using the Mu-DNA Water protocol (Sellers at al. 2018). A direct comparison between both methods revealed no evidence of bias or difference in detection probabilities (Sellers et al 2018). Field and extraction blanks were extracted alongside samples using the relevant protocol. Extraction blanks, having no filter, consisted of the reagents used in each step of the relevant protocol.

Sequencing library preparation

All samples were processed and sequenced following metabarcoding protocols established at the University of Hull using a vertebrate-specific 12S marker, amplifying a ~106 bp fragment in fish (Riaz et al. 2011; Kelly et al. 2014). Genomic DNA from non-native cichlid species (*Astatotilapia calliptera*, *Maylandia zebra* and *Rhamphochromis esox*) were used as PCR positive controls during library preparation.

Modifications to improve the molecular protocols were made between different phases of the project. In the pilot stage of the project (2015, *n* = 2 lakes), samples were PCR amplified with a one-step library preparation protocol following (Kozich et al. 2013) (see Hänfling et al. 2016a for full details). Following the pilot project, the protocol was further developed (2015–2017, *n* = 18 lakes), adopting PCR amplification using a two-step nested tagging library preparation (Kitson et al. 2019; see Li et al. 2019a; Lawson Handley at al. 2019 for full details). The most current protocol (2017 - present, *n* = 81 lakes) followed that of the nested tagging, where 24 unique tags were used for both the forward and reverse primers. Regardless of protocol, all samples were PCR amplified in triplicate, then the corresponding replicates were pooled for sequencing. For full details of the current library preparation method, see Suppl. material 1.

Bioinformatics and data set clean-up

Raw sequence data were analysed using the same bioinformatics pipeline as described in Hänfling et al. (2016a) and Li et al. (2019a). In summary, sequencing reads from all lakes underwent taxonomic assignment against a curated UK fish species reference database using a custom bioinformatics pipeline, metaBEAT (https://github.com/HullUni-bioinformatics/metaBEAT). The workflow consisted of the following steps: 1) demultiplexing; 2) trimming, quality filtering and merging; 3) chimera detection; 4) clustering; 5) taxonomic assignment. For full details of the bioinformatics workflow, see Suppl. material 1.

Following taxonomic assignment, a noise threshold of 0.1% of total reads per sample was applied to remove low frequency reads (Hänfling et al. 2016a). Most

reads were assigned to the species level, but as the molecular marker used here cannot distinguish certain species reliably, the reads belonging to these species were assigned to the next possible highest taxonomic level. Specifically, species belonging to the genera *Coregonus*, *Lampetra* and *Salvelinus* were assigned to genus level, and two members of the family Percidae (*Perca fluviatilis*, *Sander lucioperca*) were assigned to family level. Reads nominally assigned to *Lota lota* were excluded, primarily as the species is considered extinct in the UK, but also because the sequenced marker region is identical to that of the marine species *Gadus morhua*, a potential environmental contaminant via the human food chain. All remaining assignments to taxonomic levels higher than species were excluded from the analysis. Samples with fewer than 1,000 total reads or with no taxonomically assignable reads were removed. Finally, reads assigned to positive controls were removed from the data set.

Biodiversity metrics

Species richness was calculated as the total number of fish species detected within each sample (α-diversity) and across all samples for each lake (γ-diversity). Species richness estimates were calculated based on all samples of each lake and for each reduced sample number replicate to ascertain the differences between the original lake data set and that of its resampled subsets.

We also calculated two biodiversity metrics based on the relative proportion of reads for each species using "Vegan" version 2.5.6 (Oksanen et al. 2019). Simpson's reciprocal index as a measure of evenness based sample diversity and Bray-Curtis dissimilarity as a measure of difference in fish community structure between samples. As read counts from eDNA metabarcoding data have been shown to correlate strongly with actual recorded abundance and biomass of fish communities within UK freshwater systems (Li et al. 2019a; Di Muri et al. 2020) these evenness based metrics contain meaningful ecological information. However, such an inference is not an assumption of our analysis per se as we utilise evenness-based biodiversity metrics for assessing relative changes induced by sampling. This approach does not necessitate a biological interpretation of the eDNA reads.

Effect of sample number on lake fish biodiversity metrics

Three principal approaches were used to evaluate the effect of sampling effort on fish detection and community composition estimation from eDNA metabarcoding: rarefaction based species accumulation curves, statistical estimation of sampling threshold and data resampling techniques.

Species accumulation curves and statistical estimation of sampling threshold

The taxonomic completeness of sampling in individual lakes was assessed through rarefaction based species accumulation curves. Sampling threshold was calculated as the minimum number of samples required to achieve 95% of complete species richness for a given lake, which is independent of species richness and therefore comparable across different lakes. Presence/absence

data were used to determine the "sample coverage", an estimate of sample completeness, defined as the proportion of taxa in the community detected in the sample (Chao et al. 2014). Species accumulation and sample coverage were generated with "iNEXT" version 2.0.20 (Hsieh et al. 2020). We investigated the effect of lake area, species richness and lake alkalinity type on sampling threshold using Pearson rank correlations and Kruskal Wallis tests respectively.

Random resampling of lake fish eDNA metabarcoding data

A bootstrapping without replacement approach was used to generate replicate data sets with reduced sample numbers for each lake. In order to improve comparability across the data set, only lakes with ≥ 15 samples (82 lakes) were used for resampling. For each lake set consisting of *n* samples (*n* ranging from 15–20), all possible unique sample combinations at different sample sizes were generated, with sample size ranging from 2 to a maximum of *n*-2. The number of possible sample combinations drawn without replacement varies depending on total *n* and ranges from 105 (*n* = 15, 13 samples drawn) to 184,756 (*n* = 20, 10 samples drawn). For each lake, subsets of 100 unique combinations were randomly drawn and used as resampling replicates per sample size. Using this approach, there was no chance of a sample occurring more than once within a replicate, representing the reality of resampling lake samples.

The effect of sample number on species detection and community composition estimates was investigated as follows. First, the number of undetected taxa compared to the full data set was calculated for all combinations at each sample size. Here we tested for Spearman's rank coefficient correlations between the number of undetected species with total observed species richness and lake area. Values of 1, 2 and 3 were used for minimum undetected species thresholds. The sample size at which 95% of the lakes achieved less than these thresholds was considered. Second, the average deviation of a given sample combination's community composition (proportion reads) from the full lake sample composite was quantified for each sample size using pairwise dissimilarity measures (Bray-Curtis dissimilarity index). In order to quantify the effect across all lakes, the proportion of lakes which fall above an arbitrary dissimilarity value (0.1) at each sample size was calculated.

Simpson's reciprocal index was calculated using read counts per species for each lake for all combinations at each sample size and compared to the lake as a whole. The proportion variance between the values was used to gauge the level of overestimation or underestimation. All dissimilarity indices were calculated using "Vegan" version 2.5.6 (Oksanen et al. 2019)

Non-random reduced sampling of lake metabarcoding data

Random resampling provides the opportunity to explore a wide range of sample numbers but ignores the spatial context in which the samples are collected. Hence, under the assumption that eDNA is not randomly distributed, random resampling might not represent a realistic (e.g. spatially dispersed) sampling strategy. For example, with the data set analysed here, samples were collected at equidistant points around a lake perimeter. To address this, we employed a hold-out method, which better reflected the original sampling design by splitting

the samples from each lake into two interleaved subsets, i.e. two sets of 10 equidistantly distributed samples. Practically, this was achieved by grouping samples into odd and even sample numbers since samples were continuously numbered along the shoreline transect. Only lakes with exactly 20 samples (n = 63) were used for this comparison. Number of undetected species and dissimilarity indices were calculated for each lake subset as above and tested against the maximum threshold values decided for each (1 and 0.1 for undetected species and dissimilarity indices respectively). The possible effect of total species richness and lake area on the size of differences in species detection between odd and even subsets was assessed using Spearman's rank coefficient correlations.

Shoreline sampling validation

The data from shoreline and offshore samples were compared in lakes where both sample types were available (n = 20) to evaluate the generality of the findings from (Lawson Handley et al. 2019) that both sample types generate similar biodiversity estimates during the winter season.

We determined if detected species richness was affected by sample type with a linear mixed effect model. Log transformed species richness, with sample type as a covariate and lake as a random variable, was compared to the null model (no covariate of transect) with a chi-squared test of model likelihoods. Linear model analysis was performed with "lme4" version 1.1.3 (Bates et al. 2015)

Non-metric multidimensional scaling (NMDS) ordination, based on Bray-Curtis distances, was used to visualise differences in community estimates (relative abundance) between transects and the whole lake (combined transects), An analysis of similarities (ANOSIM) (Bray-Curtis dissimilarity index, 10^5 permutations) was performed to test if there were differences in relative species abundance between shoreline and offshore samples within each lake. NMDS ordination, based on Bray-Curtis distances, was used to visualise differences in relative abundance between transects. ANOSIM and NMDS was carried out using "Vegan" version 2.5.6 (Oksanen et al. 2019)

All analyses were performed using R version 4.0.5 (R Core Team 2021).

Results

Bioinformatics and data set clean-up

After taxonomic assignment, average sample read counts within lakes for each of the 101 lakes (including both shoreline and offshore samples) ranged from 13,384.30 to 101,526.60 (mean 52,646.1 ± 21,979.24 SD). Of these 2,134 samples, 2,074 remained following data set clean-up.

Effect of sample number on lake fish species biodiversity metrics

The final cleaned data set for all 101 lakes used for resampling analysis consisted of 1,792 shoreline samples. Individual lakes ranged from having 7 to 20 successfully sequenced samples with the majority (n = 63) having 20 samples. A total of 40 fish taxa were recorded across all lakes. Fish taxon richness per lake ranged from 2 to 18 (mean 7.71 ± 3.36 SD).

Species accumulation curves

Based on species accumulation estimates (Fig. 2), the majority of lakes (n = 82) had sufficient samples to detect the total species number predicted by extrapolation to 40 samples. In 10 of the remaining 19 lakes, one or more species remained undetected, and in nine lakes, two or more species remained undetected. Lakes where one or more species were potentially undetected through inadequate sampling effort tended to have higher species richness (14 of the 19 lakes had a detected species richness ≥ 10).

Sampling threshold

Regardless of actual sample size, all but five of the 101 lakes achieved sample coverage ≥ 95% for fish species detection at 20 samples (Fig. 3A), with 93 lakes achieving ≥ 95% sample coverage with a sample size of 10. A total of 96 out of 101 lakes achieved ≥ 95% sample coverage at a sample size of 11 (Fig. 3B). The sampling threshold for lakes ranged from 1 to 25 samples with the mean sample threshold at 5.37 (± 4.56 SD). Sampling threshold correlated with total species richness (r_s = 0.41, p < 0.05). There was no correlation between sampling threshold and lake surface area (r_s = -0.09, p = 0.39) or difference in sampling threshold between alkalinity types (high, medium and low) (Kruskal-Wallis: X^2 = 3.63, df = 2, p = 0.16).

Random resampling of lake metabarcoding data

The number of undetected fish species steadily decreased with increasing sample size (Fig. 4A). The point at which 95% of the lakes fall below the thresholds of 1, 2 or 3 mean species undetected were at sample sizes of 14, 9 and 6 respectively. Number of undetected species at a sample size of 10 (half the

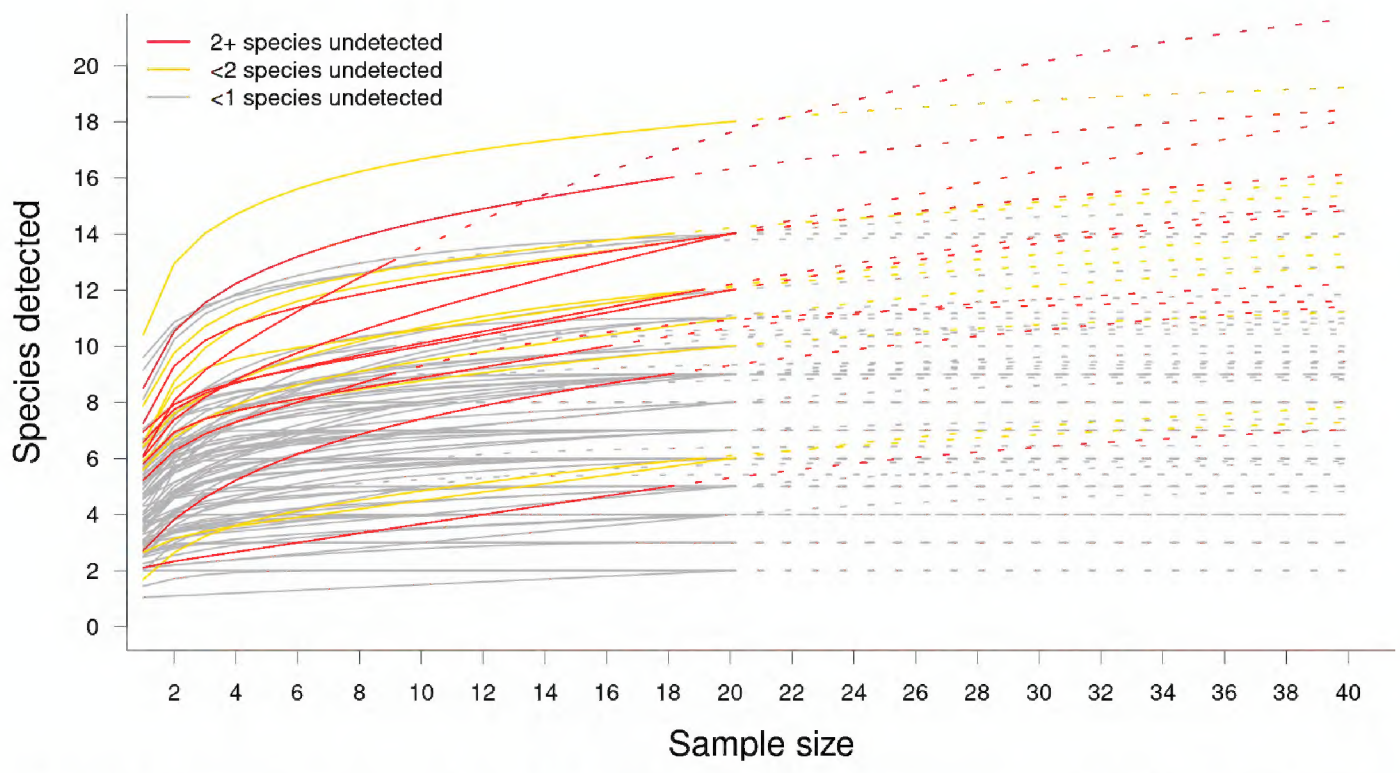

Figure 2. Species accumulation curves based on rarefaction for all 101 lakes used in this study. Grey indicates lakes with fewer than 1 estimated species undetected, yellow is lakes with fewer than 2 estimated species undetected and red is lakes with more than 2 estimated species undetected. Solid lines are interpolated, and dashed lines are extrapolated. All lakes are extrapolated to a sample size of 40 for uniformity.

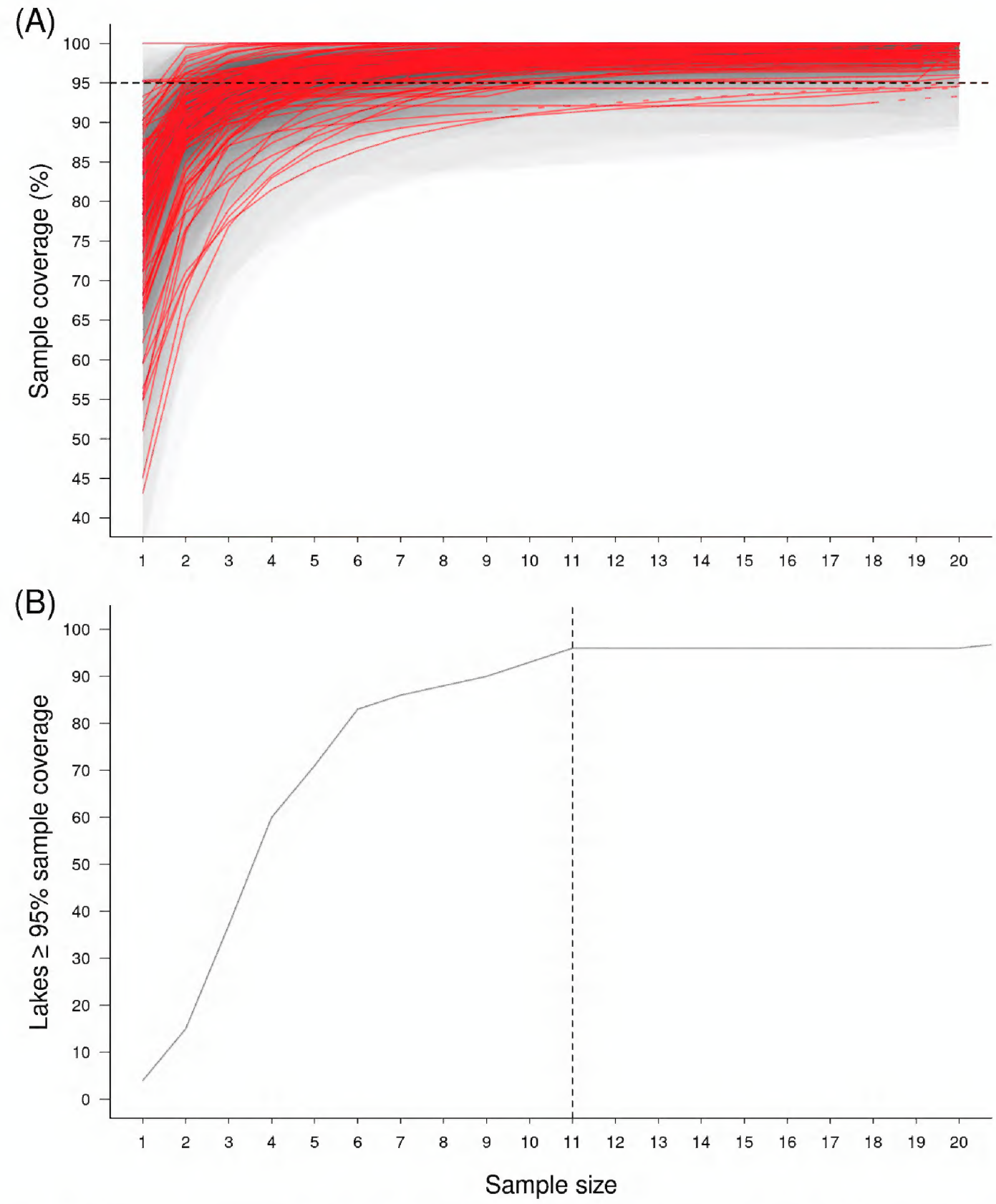

Figure 3. Sample coverage for all 101 UK lakes used in this study. Sample size cut off at 20 for uniformity **A** lake sample coverage. Solid red lines are the interpolated sample coverage. Dashed red lines are extrapolated sample coverage. Grey area shows the range of upper and lower confidence intervals. Horizontal dashed line indicates 95% sample coverage (i.e. sampling threshold) **B** cumulative count of lakes with ≥ 95% sample coverage per sample size. Vertical dashed line indicates sample size at which ≥ 95% of lakes achieve ≥ 95% sample coverage.

ideal sample size of 20 aimed for during the project) correlated with total species richness (r_s = 0.72, p < 0.05), implying that lakes with more species require a greater sampling effort for a given level of detection. There was no correlation between undetected species at sample size 10 and lake surface area (r_s = 0.07, p = 0.51). The dissimilarity index of community composition also decreased continuously with increasing sample size and ≥ 95% of the lakes fell below a mean dissimilarity index threshold of 0.1 (i.e. were more similar) at a sample size of 15 (Fig. 4B). Simpson's reciprocal index tended toward an underestimate of the lake as a whole at sample sizes less than 8 (Fig. 4C). Again, the amount of variance decreased and estimated indices became closer to the whole lake values with increased sample size.

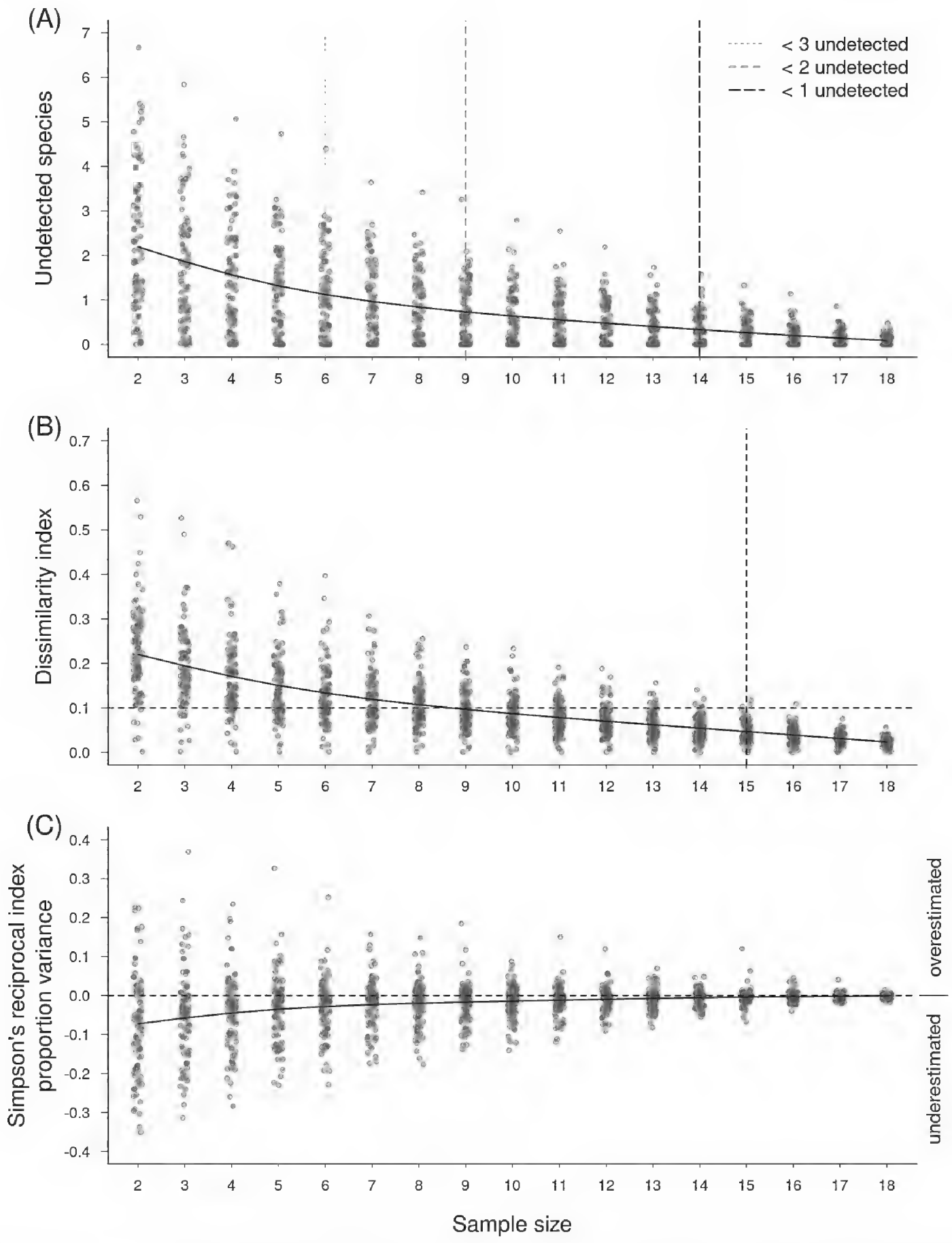

Figure 4. Random resampling of lake fish metabarcoding data from 82 lakes used in this study. All lakes analysed had a successfully sequenced sample size of ≥ 15 (maximum 20). The effects on three metrics used in the analysis are shown **A** undetected fish species counts for a lake at a given sample size. Vertical dashed lines indicate sample sizes at which ≥ 95% of lakes fell below the thresholds of 1, 2 or 3 species undetected (sample sizes of 14, 9 and 6 respectively) **B** Bray-Curtis dissimilarity index of fish communities for a lake at a given sample size to that of the whole lake. Vertical dashed line indicates sample size at which ≥ 95% of lakes achieved a mean sample dissimilarity index below an arbitrary threshold of 0.1 (horizontal dashed line) **C** proportion variance in Simpson's reciprocal index for a lake at a given sample size to that of the whole lake. In all figures, each point represents the mean of each metric for 100 unique resampling replicates of a lake at a given sample size. Solid lines show the mean of all points at a sample size.

Non-random reduced sampling of lake fish species metabarcoding data

In most cases, the number of undetected species was equal between lake subsets (*n* = 34) or differed by only a single species (*n* = 21) (Fig. 5A). In 27 of the 63 lakes, all species present were detected in both subsets. However, in a few cases (*n* = 8) the number of undetected fish species differed greatly between subsets. The size of differences in species detection between odd and even

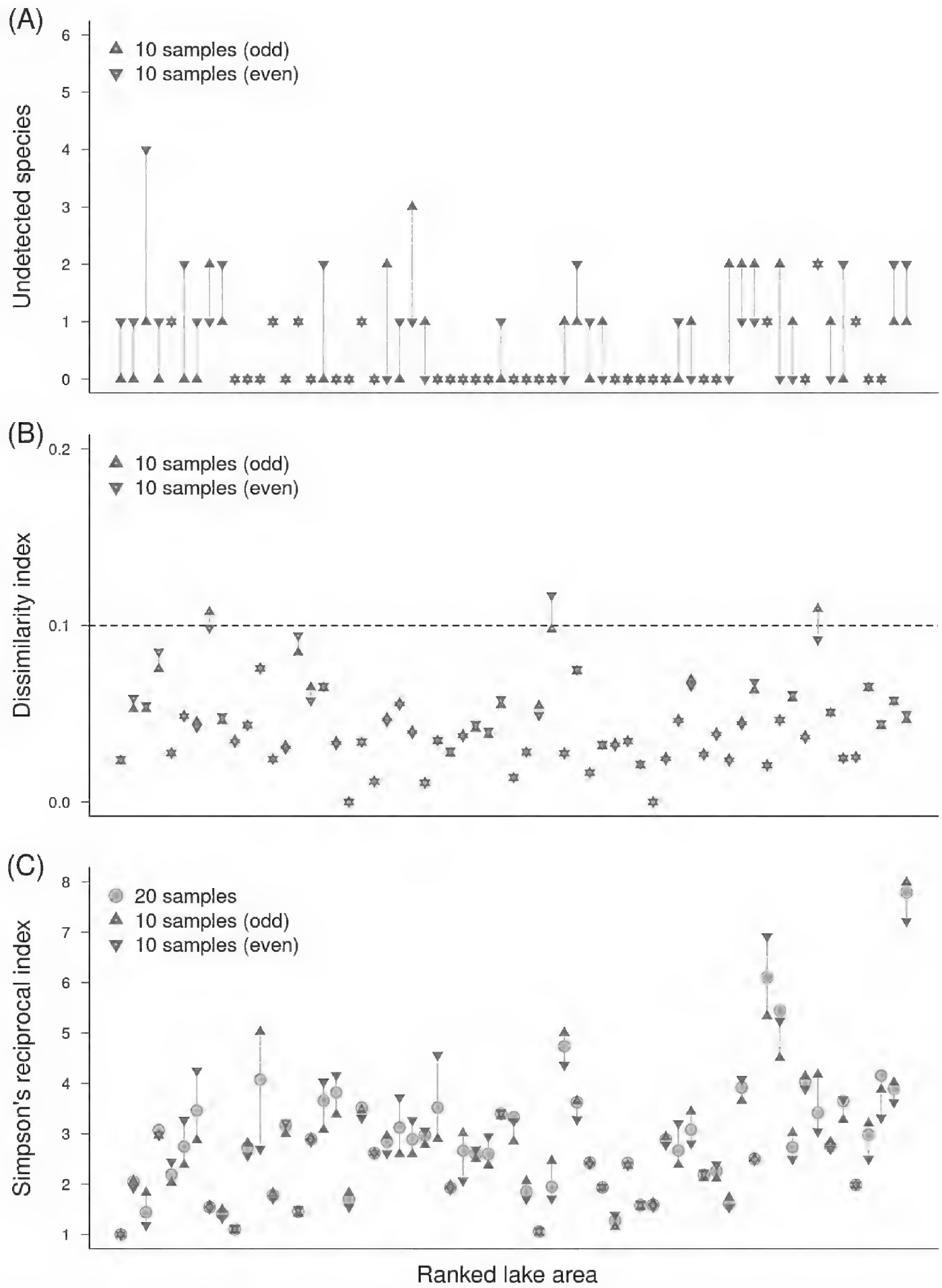

Figure 5. Non-random reduced sampling of lake fish metabarcoding data from 63 lakes used in this study. All lakes had 20 samples divided into odd (triangles) and even (inverted triangles) 10-sample subsets **A** undetected fish species counts calculated from comparison of each 10-sample subset to the whole lake **B** Bray-Curtis dissimilarity index of fish communities calculated from comparison of each subset community composition (proportion reads) to the whole lake. Horizontal dashed line indicates the decided dissimilarity index threshold (0.1) **C** Simpson's reciprocal index for odd and even subsets in comparison to the whole lake (circles). In all figures, vertical lines are visual links for corresponding lake whole, odd and even subsets. Lakes are ordered by surface area on the x-axis with size increasing from left to right.

subsets correlated with total species richness (r_s = 0.37, p < 0.05). There was no correlation with lake surface area (r_s = -0.04, p = 0.78). Differences in the Bray-Curtis dissimilarity indices of the fish communities represented in odd and even subsets per lake were generally very small and equally dissimilar to the whole lake fish community (Fig. 5B). All but three of the lakes had dissimilarity indices for both subsets below the 0.1 threshold. Simpson's reciprocal indices were highly similar for the majority of lakes with only four having more

pronounced differences between subsets and the whole lake (Fig. 5C). There was no tendency between subsets toward overestimation (odd = 31, even = 25) or underestimation (odd = 32, even = 38) of the index to that of the whole lake.

Shoreline sampling validation

A total of 34 species were present across the 20 lakes used to validate shoreline sampling, with 33 species detected in shoreline and 28 in offshore sampling transects (Fig. 6). Six species (*Alosa alosa*, *Ameiurus sp.*, *Barbus barbus*, *Blicca bjoerkna*, *Leucaspius delineatus* and *Platichthys flesus*) were unique to shoreline transects with only a single species unique to offshore transects (*Pseudorasbora parva*) (Fig. 6).

There were species unique to each transect type (i.e. shoreline and offshore) in all but one of the lakes, Loch Lubnaig (Fig. 7A). In eight of the 20 lakes, these unique species occurrences were only in shoreline samples and in 4 lakes only in the offshore samples (Fig. 7A). The majority of species detected in any given lake were shared between both transect types.

Species richness showed no significant difference between transects (X^2 = 0.121, df = 1, p = 0.728). The proportion of total species detected in transects was similar across all lakes (Fig. 7B); shoreline transects ranged from 62.5% to 100% of species detected (mean 87.36 ± 14.13 SD), and offshore from 55.65% to 100% (mean 85.43 ± 13.43 SD). With the exception of species detected only in shoreline (n = 6) or offshore (n = 1) samples, all species had similar lake occupancy scores (Fig. 6), while the exceptional species occurred

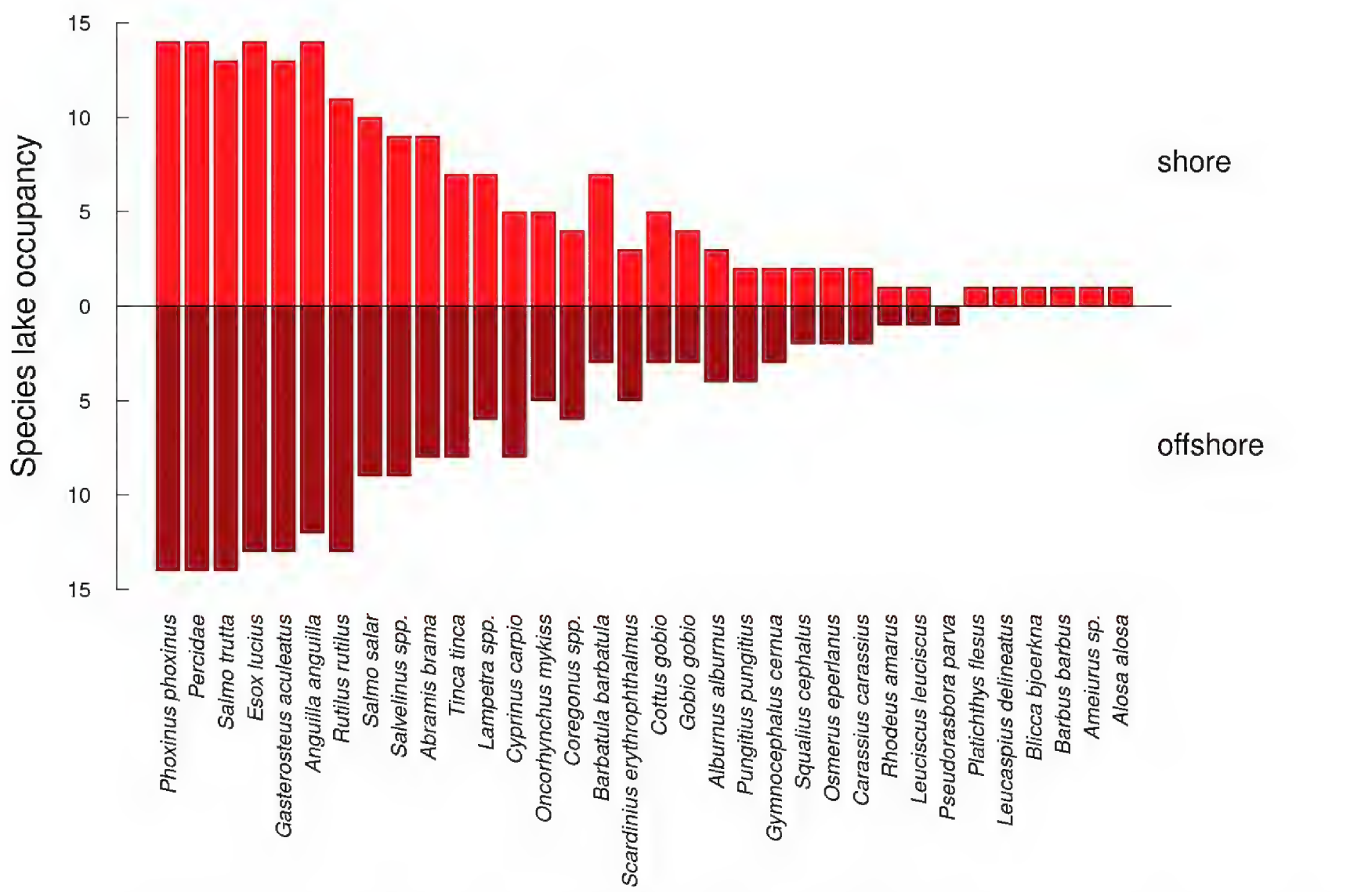

Figure 6. Species lake occupancy for shoreline and offshore sampling transects across the 20 lakes used to validate shoreline sampling. The number of lakes a species was detected in shoreline and offshore sampling transects is shown. Species are ranked by total shoreline and offshore lake occupancy.

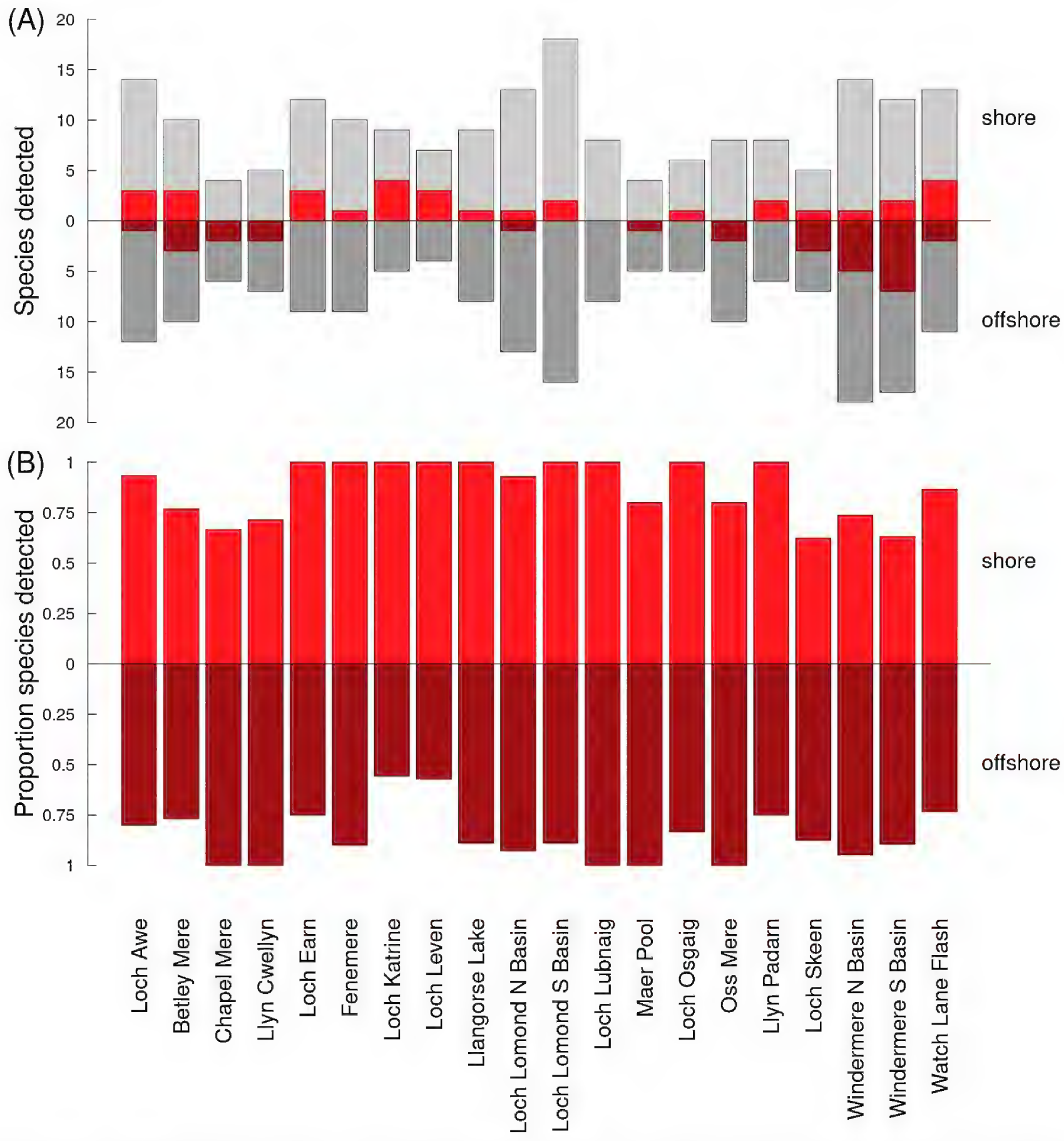

Figure 7. Overall eDNA based species detection in sampling transects of the 20 lakes used to validate shoreline sampling **A** detected species richness (grey) in shoreline and offshore sampling transects of each lake and unique species occurrences (red) for each lake **B** proportion of the total species detected using eDNA in shoreline and offshore sampling transects for each lake.

in a minority of lakes and in a minority (typically 10%) of samples from within those lakes.

Non-metric multidimensional scaling of whole lake fish community estimates (species proportion reads) demonstrated there were some differences between shoreline and offshore sampling transects (Fig. 8). However, with the exception of nine of the selected 20 lakes (those with extended ellipses), all whole lake ordinations were tightly grouped with those of their respective shoreline and offshore transects.

In contrast, on an individual lake basis, ANOSIM tests showed that there were significant differences between transect species compositions in 11 of the 20 lakes (see Suppl. material 1: fig. S1).

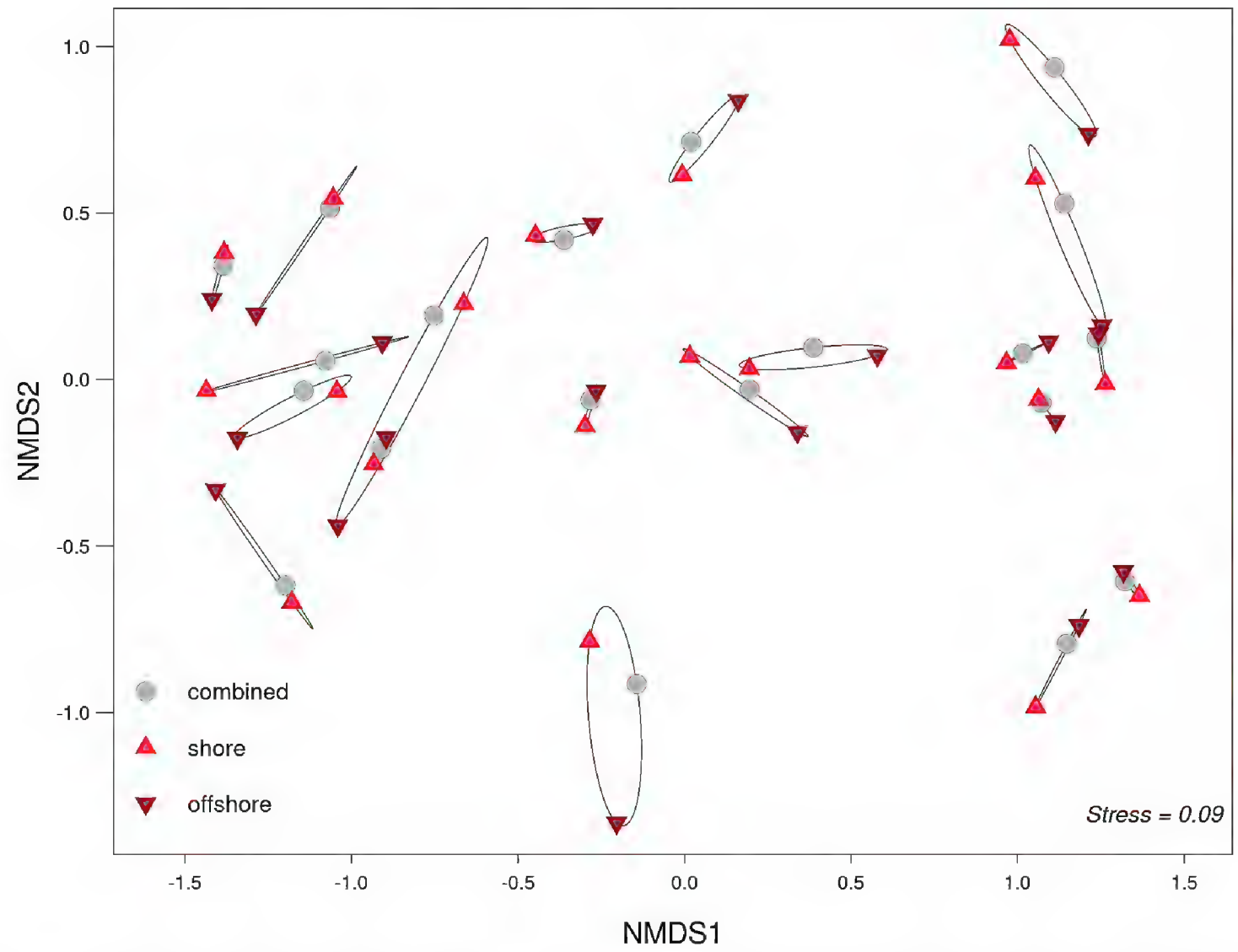

Figure 8. Non-metric multidimensional scaling (NMDS) ordination for fish communities of the 20 lakes used to validate shoreline sampling. NMDS generated from species composition (proportion reads) estimates using Bray-Curtis dissimilarity method in three dimensions (stress = 0.09). All lakes were divided into shoreline (triangles) and offshore (inverted triangles) transects. Whole lake (as both transects combined) ordinations (circles) are shown in relation to their shoreline and offshore transects. Ellipses denote the overall spread between transect composition estimates relative to that of the lake as a whole.

Discussion

This study has shown that winter shoreline sampling is an effective approach to characterise the fish community of lakes in Great Britain. The application of algorithmic and statistical resampling approaches demonstrated that 10–20 samples per lake are sufficient to detect most species and to reliably describe their relative abundance and a range of biodiversity metrics. Below we discuss the implications for designing eDNA metabarcoding surveys for lake fish communities in detail.

Effect of reduced sampling on species detection and community composition estimation

The results of the sample coverage analysis confirmed that the sampling design used to create the original data set, i.e. 20 samples from equidistant locations around the lake shore, provided a very reliable estimation of the true species richness with less than 5% of lakes (5 out of 101) having an estimated sample coverage below 95% at this sample size (Hänfling et al. 2016a; Willby et al. 2019) (Fig. 3). However, for most lakes the sample coverage curves started to reach a plateau at much lower sample numbers, indicating that the loss of signal is relatively small even with a substantially lower sampling effort. This was confirmed by the resampling analysis which indicated that in the majority of lakes, fewer than two species remain undetected on average with a sample size of 10 randomly distributed samples, and there was an even lower rate of undetected species when samples

are non-randomly distributed as would normally be the case. Interestingly, lake surface area does not directly influence the required sampling effort. However, as the required sample size increases with species richness, *a priori* knowledge of expected species richness informed by conventional sampling can be used to design efficient sampling strategies. The logistical effort of sampling is an important cost factor in eDNA-based monitoring programmes. Collection of fewer samples reduces person-hours in the field and also removes cost during downstream sample processing, such as filtration and molecular analysis.

While a reduction from 20 to 10 samples does not greatly affect ecological community analysis it does have drawbacks as the detection of locally rare or low abundance species is reduced. Therefore, sampling strategies aiming to provide accurate distribution records for species of conservation importance (e.g. endangered, or establishing invasive non-native species) which is one of the most common applications of eDNA based approaches (Piggott et al 2021; Yao et al 2022) should be based around higher sample numbers, i.e. a minimum of 20 samples per lake. The reduced sampling approach is best suited to the lower diversity lakes of Great Britain where it reliably detected the commonly occurring species making it ideal for use with established fish-based water quality assessment metrics that are not reliant on rarer species (i.e. Willby et al. 2019). Increased diversity, as is found in mainland European lakes and the rest of the world, will possibly demand an increase in sample size.

A further reduction in sample numbers could be achieved by collecting high volume samples over a transect rather than multiple point samples or at the major outflow of the lake. This is an alternative approach to the method described in this study and has been successfully employed in a number of studies to estimate species richness in lentic systems (Civade et al. 2016; Sepulveda et al. 2019; Schabacker et al. 2020) as well as large rivers (Pont et al 2018). However, this method does not provide information about the spatial distribution of species and occupancy- based abundance estimates as used in fish-based ecological quality assessment in Britain (Willby et al. 2019) and is therefore less adaptable to different project aims.

It is important to note that our results are influenced by the specific workflow used here. The detection probability of species through eDNA methods does not only depend on the number of samples taken within a habitat, but also on levels of replication during other stages of the workflow such as PCR and sequencing (Ficetola et al. 2015). Furthermore, the specific laboratory protocols such as the choice of extraction method, choice of primer, number of amplification cycles or TaqPolymerase could also affect detection probability. Hence findings may differ if methods are used which have lower or higher detection probabilities within individual samples. For example, fewer samples than indicated in our study might be needed if more than three PCR replicates per sample are used. However, it is likely that the broad trends we detected will be similar irrespective of such changes.

Spatio-temporal considerations of sampling

Our extensive resampling analysis of eDNA metabarcoding data collected from the shore of more than 100 lakes during the winter season showed that utilising 10–20 samples was sufficient for detecting most fish species present in a

typical lake in Great Britain. Moreover, within a smaller subset of lakes (n = 20) which included some of the UKs largest lakes that had both shoreline and offshore transect samples, we observed no differences in species diversity (i.e. number of species detected) between offshore and shoreline samples. These results strongly support the effectiveness of winter shoreline sampling as a reliable method for fish species detection in lakes of Great Britain. This conclusion is in line with previous research conducted in Windermere, England (Lawson Handley et al. 2019) and three Chinese lakes which were sampled during the autumn (Zhang et al. 2020). One contributing factor to the success of winter shoreline sampling might be that the specific hydrological processes affecting temperate are lacking during autumn and winter seasons. Increased water circulation due to the absence of thermal stratification, facilitates eDNA dispersal from the deeper areas of the lake to the shore. Additionally, the low temperatures during these seasons can slow down DNA degradation processes (Jo et al. 2019; Harrison et al. 2019). Further support for this comes from a study in three French lakes which also demonstrated that offshore sampling was unnecessary when lakes lacked stratification (Hervé et al. 2022). In contrast, DNA dispersal might be more limited during warmer seasons. (Littlefair et al. 2021) showed that stratification of Canadian lakes prevented detection of deepwater species throughout the water column. Our investigation in Windermere also revealed a more localised distribution of eDNA during the summer, with fewer species detected in shoreline samples compared to winter (Lawson Handley et al. 2019). Additionally, studies on the spatial distribution of eDNA in summer ponds using cage experiments have shown a drastic decreases in eDNA detection probability after distances of 5–10 m from the source (Li et al. 2019a; Brys et al. 2021). Collectively, this evidence suggests that a sampling strategy based exclusively on shoreline sampling is effective during polymictic conditions in autumn and/or winter, but may be less effective during the summer months. As sample site access is a major logistical concern and shoreline sites are generally more accessible than offshore sites, removing the potential complications of boat use to access offshore sites would be highly beneficial for lake monitoring. Even in lakes with difficult land access to the shoreline, boat sampling of surface water near the shoreline is logistically easier than collecting samples in deeper water offshore that requires more specialised water sampling equipment. These simpler logistics suggested by our results therefore further help to reduce the costs of lake eDNA sampling programmes. For example, pelagic/profundal offshore species such as Coregonus and S. alpinus were detected by winter shoreline sampling.

While there was no evidence of a difference in detection probability between shoreline and offshore samples for any individual species across the entire data set, the species composition differed significantly between offshore and shoreline samples in 11 out of 20 lakes. However, these differences were relatively small compared to differences among lakes and mainly due to variation in relative abundance of some frequent species. Some rare species were only present in one of the two sample types. This is likely due to stochastic effects as there was no evidence of a systematic bias for individual species in relation to transect type across the data set (Fig. 6). These exceptional species were also rare within the lakes where they were found. Nevertheless, monitoring programmes need to consider potential

differences between offshore and shoreline samples when measuring temporal trends in community composition and use a consistent sampling approach over time.

In the data set analysed here, we detected some fish species more typically associated with river systems (rheophilic fish) in lake water samples, such as European bullhead (*Cottus gobio*), grayling (*Thymallus thymallus*), lamprey (Lampetra spp.) and salmon (*Salmo salar*). Rivers have been shown to transport eDNA over great distances (Deiner et al. 2016), although eDNA quantity decreases rapidly during this process (Pont et al. 2018). Hence some detections, especially rare ones, could reflect influence from upstream river water. However, rheophilic fish also occur in lake estuaries, stray into the lakes or utilise lakes for a part of their life cycle (e.g. salmonids (Arostegui and Quinn 2019)). From sequencing data alone, it is therefore impossible to disentangle if detection within a lake is true occupancy or transport of eDNA from upstream rivers. It is therefore more appropriate to regard the eDNA sampling in lakes as sampling of the lake itself and locally connected freshwater habitat.

Conclusion

The results of this study provide an important overview of how sampling effort and design affect various metrics of fish species richness in lakes which will provide guidance on optimising sampling strategies for individual projects. This will, however, require projects to have clear objectives and predefined standards in terms of acceptable error. As a general rule, to achieve an overview of species composition in relatively low fish diversity lakes, as is typical for many regions of Great Britain, 10 samples per lake taken during the winter season will suffice, regardless of lake surface area. However, sample size will need to be increased if detection of rarer species is required or is a priority, or when sampling high diversity lakes. These results are not necessarily directly transferable to other systems as different temperature regimes and hydrological conditions are likely to affect the spatial distribution and detection probability of eDNA in lentic systems. Although our understanding of these factors has improved considerably over the last ten years, there is still a knowledge gap in the effect of seasonal variation in detection in different ecosystems. The approach presented here should be seen as a framework for optimising sampling effort in other lentic ecosystems.

Acknowledgments

The majority of the data used for this study are owned by the Environment Agency, Natural Resources Wales and Scottish Environment Protection Agency. Natural England funded data analysis used to inform the content of this paper. Without exceptions, land owners and fishery owners have supported data collection by allowing access to the relevant water bodies.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

No funding was reported.

Author contributions

Conception and study design: BH, GS, DO. Sample collection and field work: LRH, MB, CDM, JL, LLH, BH. Site selection and lake metadata: GP, KW, THE, WD, AD, CWB, IJW. Laboratory work and bioinformatics: LRH, MB, CDM, JL, GS. Data analysis: GS, CLJ, NW, AL. Data curation: LRH, CDM, GS. Writing - original draft: BH, GS. All authors contributed to reviewing and editing of the manuscript.

Author ORCIDs

Graham S. Sellers https://orcid.org/0000-0002-9679-0435
Christopher L. Jerde https://orcid.org/0000-0002-8074-3466
Lynsey R. Harper https://orcid.org/0000-0003-0923-1801
Marco Benucci https://orcid.org/0000-0003-2463-1467
Cristina Di Muri https://orcid.org/0000-0003-4072-0662
Jianlong Li https://orcid.org/0000-0002-0302-0061
Dave Ottewell https://orcid.org/0009-0004-0411-8038
Alan Law https://orcid.org/0000-0001-5971-3214
Colin W. Bean https://orcid.org/0000-0003-3502-0995
Ian J. Winfield https://orcid.org/0000-0001-9296-5114

Data availability

Newly generated sequencing data have been deposited as raw demultiplexed reads in NCBI's Sequence Read Archive (SRA) under project numbers PRJNA1032436 and PRJNA454866 and are publicly accessible. Previously published data from Li et al. (2019a) can be found under project PRJNA454866. All data supporting the findings of this study, along with associated analysis scripts, are openly available in the Supporting Information at http://dx.doi.org/10.17605/OSF.IO/WVCKM.

References

Arostegui MC, Quinn TP (2019) Reliance on lakes by salmon, trout and charr (*Oncorhynchus*, *Salmo* and *Salvelinus*): An evaluation of spawning habitats, rearing strategies and trophic polymorphisms. Fish and Fisheries 20(4): 775–794. https://doi.org/10.1111/faf.12377

Barnes M, Turner C (2015) The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics 17: 1–17. https://doi.org/10.1007/s10592-015-0775-4

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67(1): 1–48. https://doi.org/10.18637/jss.v067.i01

Bedwell ME, Goldberg CS (2020) Spatial and temporal patterns of environmental DNA detection to inform sampling protocols in lentic and lotic osystems. Ecology and Evolution 10(3): 1602–1612. https://doi.org/10.1002/ece3.6014

Beentjes KK, Speksnijder AGCL, Schilthuizen M, Hoogeveen M, van der Hoorn BB (2019) The effects of spatial and temporal replicate sampling on eDNA metabarcoding. PeerJ 7: e7335. https://doi.org/10.7717/peerj.7335

Birk S, Willby NJ, Kelly MG, Bonne W, Borja A, Poikane S, van de Bund W (2013) Intercalibrating classifications of ecological status: Europe's quest for common management objectives for aquatic ecosystems. Science of the Total Environment 454–455: 490–499. https://doi.org/10.1016/j.scitotenv.2013.03.037

Blackman RC, Osathanunkul M, Brantschen J, Di Muri C, Harper LR, Mächler E, Hänfling B, Altermatt F (2021) Mapping biodiversity hotspots of fish communities in subtropical streams through environmental DNA. Scientific Reports 11(1): 10375. https://doi.org/10.1038/s41598-021-89942-6

Bruce K, Blackman RC, Bourlat SJ, Hellström M (2021) A practical guide to DNA-based methods for biodiversity assessment. Advanced Books. https://doi.org/10.3897/ab.e68634

Brys R, Haegeman A, Halfmaerten D, Neyrinck S, Staelens A, Auwerx J, Ruttink T (2021) Monitoring of spatiotemporal occupancy patterns of fish and amphibian species in a lentic aquatic system using environmental DNA. Molecular Ecology 30(13): 3097–3110. https://doi.org/10.1111/mec.15742

Cantera I, Cilleros K, Valentini A, Cerdan A, Dejean T, Iribar A, Taberlet P, Vigouroux R, Brosse S (2019) Optimizing environmental DNA sampling effort for fish inventories in tropical streams and rivers. Scientific Reports 9(1): 3085. https://doi.org/10.1038/s41598-019-39399-5

Chao A, Gotelli NJ, Hsie TC, Sander EL, Ma KH, Colwell RK, Ellison AM (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecological Monographs 84: 45–67. https://doi.org/10.1890/13-0133.1

Civade R, Dejean T, Valentini A, Roset N, Raymond J-C, Bonin A, Taberlet P, Pont D (2016) Spatial Representativeness of Environmental DNA Metabarcoding Signal for Fish Biodiversity Assessment in a Natural Freshwater System. PLoS ONE 11: e0157366. https://doi.org/10.1371/journal.pone.0157366

Czeglédi I, Sály P, Specziár A, Preiszner B, Szalóky Z, Maroda Á, Pont D, Meulenbroek P, Valentini A, Erős T (2021) Congruency between two traditional and eDNA-based sampling methods in characterising taxonomic and trait-based structure of fish communities and community-environment relationships in lentic environment. Ecological Indicators 129: 107952. https://doi.org/10.1016/j.ecolind.2021.107952

Deiner K, Fronhofer EA, Mächler E, Walser J-C, Altermatt F (2016) Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature Communications 7(1): 12544. https://doi.org/10.1038/ncomms12544

Dejean T, Valentini A, Miquel C, Taberlet P, Bellemain E, Miaud C (2012) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog Lithobates catesbeianus: Alien invasive species detection using eDNA. Journal of Applied Ecology 49(4): 953–959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Di Muri C, Lawson Handley L, Bean CW, Li J, Peirson G, Sellers GS, Walsh K, Watson HV, Winfield IJ, Hänfling B (2020) Read counts from environmental DNA (eDNA) metabarcoding reflect fish abundance and biomass in drained ponds. Metabarcoding and Metagenomics 4: e56959. https://doi.org/10.3897/mbmg.4.56959

Dunker KJ, Sepulveda AJ, Massengill RL, Olsen JB, Russ OL, Wenburg JK, Antonovich A (2016) Potential of Environmental DNA to Evaluate Northern Pike (Esox lucius) Eradication Efforts: An Experimental Test and Case Study. PLoS ONE 11(9): e016227. https://doi.org/10.1371/journal.pone.0162277

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources 15(3): 543–556. https://doi.org/10.1111/1755-0998.12338

Griffiths NP, Bolland JD, Wright RM, Murphy LA, Donnelly RK, Watson HV, Hänfling B (2020) Environmental DNA metabarcoding provides enhanced detection of the European eel Anguilla anguilla and fish community structure in pumped river catchments. Journal of Fish Biology 97(5): 1375–1384. https://doi.org/10.1111/jfb.14497

Hänfling B, Lawson Handley L, Read DS, Winfield IJ (2016b) The development of an eDNA-based approach for fish sampling in lochs for WFD - Phase 2. Report to the Scottish Environment Protection Agency, UK.

Hänfling B, Lawson Handley L, Read DS, Winfield IJ (2016c) eDNA-based metabarcoding as a monitoring tool for fish in large lakes. Report – SC140018/R. Report to the Environment Agency, UK.

Hänfling B, Lawson Handley L, Harper LR, Benucci M, Sellers GS, Di Muri C, Griffiths N, Jaques R, James J (2020) Development of an eDNA-based lake fish classification tool – Collection of data from English Lakes. Report to the Environment Agency, UK. [Unpublished]

Hänfling B, Lawson Handley L, Read DS, Hahn C, Li J, Nichols P, Blackman RC, Oliver A, Winfield IJ (2016a) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25(13): 3101–3119. https://doi.org/10.1111/mec.13660

Harrison JB, Sunday JM, Rogers SM (2019) Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings. Biological Sciences 286(1915): 20191409. https://doi.org/10.1098/rspb.2019.1409

Hayami K, Sakata MK, Inagawa T, Okitsu J, Katano I, Doi H, Nakai K, Ichiyanagi H, Gotoh RO, Miya M, Sato H, Yamanaka H, Minamoto T (2020) Effects of sampling seasons and locations on fish environmental DNA metabarcoding in dam reservoirs. Ecology and Evolution 10(12): 5354–5367. https://doi.org/10.1002/ece3.6279

Hervé A, Domaizon I, Baudoin J-M, Dejean T, Gibert P, Jean P, Peroux T, Raymond J-C, Valentini A, Vautier M, Logez M (2022) Spatio-temporal variability of eDNA signal and its implication for fish monitoring in lakes. PLoS ONE 17(8): e0272660. https://doi.org/10.1371/journal.pone.0272660

Hsieh TC, Ma KA, Chao A (2020) iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.20. http://chao.stat.nthu.edu.tw/wordpress/software-download/

Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA: eDNA surveillance of rare aquatic species. Conservation Letters 4(2): 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jo T, Murakami H, Yamamoto S, Masuda R, Minamoto T (2019) Effect of water temperature and fish biomass on environmental DNA shedding, degradation, and size distribution. Ecology and Evolution 9(3): 1135–1146. https://doi.org/10.1002/ece3.4802

Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using environmental DNA to census marine fishes in a large mesocosm. PLoS ONE 9(1): e86175. https://doi.org/10.1371/journal.pone.0086175

Kitson JJN, Hahn C, Sands RJ, Straw NA, Evans DM, Lunt DH (2019) Detecting host-parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. Molecular Ecology 28(2): 471–483. https://doi.org/10.1111/mec.14518

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology 79(17): 5112–5120. https://doi.org/10.1128/AEM.01043-13

Lawson Handley L, Read DS, Winfield IJ, Kimbell H, Johnson H, Li J, Hahn C, Blackman R, Wilcox R, Donnelly R, Szitenberg A, Hänfling B (2019) Temporal and spatial variation in distribution of fish environmental DNA in England's largest lake. Environmental DNA 1(1): 26–39. https://doi.org/10.1002/edn3.5

Li J, Lawson Handley L-J, Read DS, Hänfling B (2018) The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. Molecular Ecology Resources 18(5): 1102–1114. https://doi.org/10.1111/1755-0998.12899

Li J, Hatton-Ellis TW, Lawson Handley L, Kimbell HS, Benucci M, Peirson G, Hänfling B (2019a) Ground-truthing of a fish-based environmental DNA metabarcoding method for assessing the quality of lakes. Journal of Applied Ecology 56(5): 1232–1244. https://doi.org/10.1111/1365-2664.13352

Li J, Lawson Handley LJ, Harper LR, Brys R (2019b) Limited dispersion and quick degradation of environmental DNA in fish ponds inferred by metabarcoding. Environmental DNA 1: 238–250. https://doi.org/10.1002/edn3.24

Littlefair JE, Hrenchuk LE, Blanchfield PJ, Rennie MD, Cristescu ME (2021) Thermal stratification and fish thermal preference explain vertical eDNA distributions in lakes. Molecular Ecology 30(13): 3083–3096. https://doi.org/10.1111/mec.15623

McElroy ME, Dressler TL, Titcomb GC, Wilson EA, Deiner K, Dudley TL, Eliason EJ, Evans NT, Gaines SD, Lafferty KD, Lamberti GA, Li Y, Lodge DM, Love MS, Mahon AR, Pfrender ME, Renshaw MA, Selkoe KA, Jerde CL (2020) Calibrating environmental DNA metabarcoding to conventional surveys for measuring fish species richness. Frontiers in Ecology and Evolution 8: 276. https://doi.org/10.3389/fevo.2020.00276

Milián-García Y, Young R, Madden M, Bullas-Appleton E, Hanner RH (2021) Optimization and validation of a cost-effective protocol for biosurveillance of invasive alien species. Ecology and Evolution 11(5): 1999–2014. https://doi.org/10.1002/ece3.7139

Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHM, Szoecs E, Wagner H (2019) vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Piggott MP, Banks SC, Broadhurst BT, Fulton CJ, Lintermans M (2021) Comparison of traditional and environmental DNA survey methods for detecting rare and abundant freshwater fish. Aquatic Conservation 31(1): 173–184. https://doi.org/10.1002/aqc.3474

Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N, Schabuss M, Zornig H, Dejean T (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports 8(1): 10361. https://doi.org/10.1038/s41598-018-28424-8

Pukk L, Kanefsky J, Heathman AL, Weise EM, Nathan LR, Herbst SJ, Sard NM, Scribner KT, Robinson JD (2021) eDNA metabarcoding in lakes to quantify influences of landscape features and human activity on aquatic invasive species prevalence and fish community diversity. Diversity & Distributions 27(10): 2016–2031. https://doi.org/10.1111/ddi.13370

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research 39(21): e145. https://doi.org/10.1093/nar/gkr732

Sato H, Sogo Y, Doi H, Yamanaka H (2017) Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports 7(1): 14860. https://doi.org/10.1038/s41598-017-14978-6

Schabacker JC, Amish SJ, Ellis BK, Gardner B, Miller DL, Rutledge EA, Sepulveda AJ, Luikart G (2020) Increased eDNA detection sensitivity using a novel high-volume water sampling method. Environmental DNA 2(2): 244–251. https://doi.org/10.1002/edn3.63

Sellers GS, Di Muri C, Gómez A, Hänfling B (2018) Mu-DNA: A modular universal DNA extraction method adaptable for a wide range of sample types. Metabarcoding and Metagenomics 2: e24556. https://doi.org/10.3897/mbmg.2.24556

Sepulveda AJ, Schabacker J, Smith S, Al-Chokhachy R, Luikart G, Amish SJ (2019) Improved detection of rare, endangered and invasive trout in using a new large-volume sampling method for eDNA capture. Environmental DNA 1(3): 227–237. https://doi.org/10.1002/edn3.23

UKTAG [UK Technical Advisory Group on the Water Framework Directive] (2004) Guidance on Typology for Lakes for the UK. Water Framework Directive (WFD). https://www.wfduk.org/sites/default/files/Media/Characterisation%20of%20the%20water%20environment/Lakes%20typology_Final_010604.pdf

Wang S, Yan Z, Hänfling B, Zheng X, Wang P, Fan J, Li J (2021) Methodology of fish eDNA and its applications in ecology and environment. The Science of the Total Environment 755: 142622. https://doi.org/10.1016/j.scitotenv.2020.142622

Willby N, Law A, Bull C, Hänfling B, Lawson Handley L, Winfield IJ (2019) A tool for classifying the ecological status of lake fish in Britain based on eDNA metabarcoding. Report to the Scottish Environment Protection Agency, UK.

Yao M, Zhang S, Lu Q, Chen X, Zhang S-Y, Kong Y, Zhao J (2022) Fishing for fish environmental DNA: Ecological applications, methodological considerations, surveying designs, and ways forward. Molecular Ecology 31(20): 5132–5164. https://doi.org/10.1111/mec.16659

Zhang S, Lu Q, Wang Y, Wang X, Zhao J, Yao M (2020) Assessment of fish communities using environmental DNA: Effect of spatial sampling design in lentic systems of different sizes. Molecular Ecology Resources 20(1): 242–255. https://doi.org/10.1111/1755-0998.13105

Supplementary material 1

Supporting information

Authors: Graham S. Sellers, Christopher L. Jerde, Lynsey R. Harper, Marco Benucci, Cristina Di Muri, Jianlong Li, Graeme Peirson, Kerry Walsh, Tristan Hatton-Ellis, Willie Duncan, Alistair Duguid, Dave Ottewell, Nigel Willby, Alan Law, Colin W. Bean, Ian J. Winfield, Daniel S. Read, Lori Lawson Handley, Bernd Hänfling

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.8.104655.suppl1